The Effect of Categorization and Public Transportation on Food Deserts in Cook County, IL

Lily Boqing Xu
June 12, 2015
Advisor: Professor Bruce Spencer
Acknowledgements

First and foremost, I would like to take this opportunity to thank my advisor, Professor Bruce Spencer, for his inspiration on this thesis topic and for his invaluable instruction and guidance throughout this past year. His patience and support encouraged me to overcome the various challenges I encountered during the thesis research and writing processes. I would also like to thank Professor Joseph Ferrie for sharing his knowledge on food deserts, and for his direction and understanding throughout the writing of this thesis. And last but not least, I would like to thank Professor William Rogerson and Sarah M. Ferrer for all of their advice and assistance over my four years as part of MMSS.
Abstract

The problem with the current definition of food deserts provided by the U.S. Department of Agriculture (USDA) is that it only has a minimum threshold requirement and does not divide food deserts into differing levels of severity in terms of food access. Cook County has 36 identified food deserts that need help. I first reassess the current state of the 36 identified food deserts based off of the USDA definition, using correlation and regression analyses to estimate the relationship between low income percentage and grocery store availability and accessibility. I find that not all of these food deserts are equal in severity. Unfortunately, because I am unable to find previous literature that divides food deserts into categorizes, I create my own criteria (based off of the original USDA threshold) to classify the food deserts into four groups of severity levels of food access. I then compare these levels of the food desert census tracts after incorporating public transportation into the definition. With the current financial state of the City of Chicago and Cook County, resources are limited. In order to best allocate these limited resources to generate the most impact on improving healthy food accessibility, I identify the food deserts in Cook County that requires the most immediate attention through this new definition. Lastly, I propose possible transportation-related solutions to tackle the issues of healthy food accessibility.
1. Introduction

1.1 What are food deserts?

According to the USDA, Treasury, and HHS, a food desert is defined as a "census tract with a substantial share of residents who live in low-income areas that have low levels of access to a grocery store or healthy, affordable food retail outlet."

A census tract is considered a **low-income community** if it has either 1) a poverty rate of 20% or greater OR 2) a median family income at or below 80% of the area median family income.

A census tract is considered a **low-access community** if at least 33% of the tract's population, or a minimum of 500 people in the tract, live more than one mile from a supermarket or large grocery store. This distance is measured by the distance between the geographic center of the census tract and the nearest large grocery store.

For a census tract to qualify as a food desert by USDA standards, they must meet both the low-income and low-access thresholds.

1.2 What is so important about food deserts?

According to Mari Gallagher, a pioneer in the study of food deserts, particularly in Chicago, and the founder of the national Mari Gallagher Research and Consulting Group:

“A food desert is the antithesis of progress, and the costs associated with living within one will be borne directly by those residents through their quality and length of life, and indirectly by the health care industry, by employers, by government agencies, and by others who take of the financial burden of pre-death treatments.”

With the lack of grocery store availability and easily accessible public transportation, many residents are denied the access to fresh produce that all people are entitled to. In most
urban areas, people residing in lower income and minority communities face barriers that limit their access to healthy food (Kim, D). Due to the limited number of stores, supermarkets may not be readily accessible. And thus, mobility to the already limited supermarkets becomes an even more important factor when addressing food accessibility (Kim, D).

The USDA’s Economic Research Service estimates that 23.5 million people in the continental United States live in food desert census tracts. Over half of that number, 23.5 million people, are considered low income. About 2.3 million people live more than a mile away from a supermarket and do not own a car. More specifically in Chicago, more than 500,000 residents live in food deserts. Below is an illustration of where these 500,000 food deserts Chicagoans live.

**Figure 1. Food Desert Map of Chicago Neighborhoods**

1.3 Previous research on transportation-related solutions

"Food accessibility in underserved areas, in terms of spatiality and mobility, is overshadowed by the incentives to provide greater availability of foods" (Kim, D). Availability addresses the idea to provide greater fresh food options and have them readily available in certain distances (Nikhanj, S). However, grocery store availability fails to address the accessibility barriers residents of food deserts face, such as long walking distances and limited public transportation options. And thus, the problem with accessibility "includes availability and the determination of resources needed to get to healthy food options in afflicted communities" (Kim, D). Many residents in food deserts not only struggle with the low availability of nearby healthy food options, but also with the low accessibility to these grocery store locations.

Unfortunately, most public transportation systems do not consider convenient access to grocery stores when designing their routes. In fact, bus lines rarely are designed to serve intra-neighborhood food shopping patterns and usually planned around commuter routes (Gottlieb, R). Due to this restriction in transportation options, inner city residents have less access to fresh produce (Gottlieb, R). This lack of direct access to grocery stores create more obstacles for residents, as they have the extra burden of carrying grocery bags from supermarkets back home.

"Since transportation barriers and long distances comes into play, many may opt to choose the most convenient and closest alternatives" (Kim, D). In low access areas, the only available food option may be at the local convenience store, which most likely does not carry healthy options, let alone fresh fruits and vegetables. As a result, walking distance becomes a important factor in food purchase behavior, which can result in inadequate and unhealthy food choices.
2. Purpose of Thesis

The purpose of my thesis is threefold. First, I will assess the state of Cook County's currently identified food desert areas based on the USDA definition of low access. Second, I will reevaluate the severity of Cook County's food deserts after incorporating public transportation into the 1 mile walking distance threshold that qualifies a census tract as a food desert. And last, I wish to identify the most severe food desert census tracts and provide possible solutions. For this thesis, I will refer to a food desert having greater severity if a resident has to walk further to reach the nearest large grocery store from the geographic center of the census tract.

3. Data

All information used in this thesis is based off of the 2010 Census by the U.S. Census Bureau. The identification of Cook County food desert census tracts and the demographic information of residents in each census tract is extracted from the 2010 Archived Food Access Research Atlas Data from the USDA Economic Research Service. The remaining data about each census tract are self-collected from USBoundary.com and Google Maps.

3.1 Self-Collected Data on Accessibility within Each Census Tract

The definition of food deserts provided by the USDA measures the walking distance between the geographic center of a census tract to the nearest large grocery store. Unfortunately, because this information is not readily available, I create my own variables to measure the accessibility problem, both with and without the use of public transportation, for each identified food desert in Cook County.

After inputting the census tract number from the Food Access Research Atlas Data into USBoundary.com, I collect the following:

- The land area of census tract in square miles
- Total number of households and housing units
- The latitude and longitude of the geographic center of the census tract

I then enter the coordinates of the center into Google Maps and use the Search Nearby option to locate the "nearest grocery store." With this feature, I count both the number of grocery stores and the number of bus stops within the census tract boundaries. After finding all of the grocery stores surrounding the center of the census tract, I record the one that requires the shortest walking distance (indicated on Figure 2 by the red arrow).

**Figure 2.** Identifying the nearest grocery store to the geographic center of census tract

From there, I look up both the walking and public transportation directions (Figure 3) from the geographic center to the grocery store and note the following:

- Walking distance from the center of the census tract to the nearest grocery store
- Walking distance from the center of the census tract to the nearest bus stop to get to the grocery store via public transportation
- Walking distance from the grocery store to its nearest bus stop
• Total walking distance from the center to the nearest grocery store with the aid of public transportation

By using Google Maps, I can see the current status on grocery store availability and accessibility and find if there have been any improvements in the food deserts since being identified five years ago.

Figure 3. Walking distance to nearest grocery store with and without use of public transportation

3.3 Data Description

Below are the names and definition of the key variables I utilize for this thesis.

• Tract_FIPS = Federal Information Processing (FIPS) Tract Identifier
- PERCENT_LOWI = Percentage of total population that is low-income and has low access to a supermarket or large grocery store
- Wdist(CG) = Walking distance from center of census tract to nearest grocery store
- Wdist CB = Walking distance from center of census tract to nearest bus stop
- Wdist BG = Walking distance from nearest grocery store to its nearest bus stop
- Wdist TotB = Wdist CB + Wdist BG
  = Total walking distance to nearest grocery store after taking the bus
- Land Area = Land area of census tract in square miles
- TotalHH = Total number of households
- NumBusStops = Number of bus stops within census tract
- NumBusStopsPerArea = Number of bus stops within census tract per land area
- NumGroceryStores = Number of grocery stores within census tract
- NumGroceryPerArea = Number of grocery stores in census tract per land area
- NumGroceryPer100HH = Number of grocery stores in census tract per 100 households

Table 1. Summary of Key Variables Used in Analysis

<table>
<thead>
<tr>
<th>Variable</th>
<th>Label</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tract_FIPS</td>
<td>Tract_FIPS</td>
<td>36</td>
<td>1703171387</td>
<td>145104.61</td>
<td>17031420100</td>
<td>17031830001</td>
</tr>
<tr>
<td>Wdist_TotB</td>
<td>Wdist_TotB</td>
<td>36</td>
<td>0.5431581</td>
<td>0.5369162</td>
<td>0.0634470</td>
<td>2.6732955</td>
</tr>
<tr>
<td>Wdist(CG)</td>
<td>Wdist(CG)</td>
<td>36</td>
<td>0.8546507</td>
<td>0.7592742</td>
<td>0.0174242</td>
<td>3.2000000</td>
</tr>
<tr>
<td>Wdist BG</td>
<td>Wdist BG</td>
<td>36</td>
<td>0.1324258</td>
<td>0.193956</td>
<td>0</td>
<td>1.0000000</td>
</tr>
<tr>
<td>Wdist CB</td>
<td>Wdist CB</td>
<td>36</td>
<td>0.4107323</td>
<td>0.5063802</td>
<td>0</td>
<td>2.6600000</td>
</tr>
<tr>
<td>NumBusStops</td>
<td>NumBusStops</td>
<td>36</td>
<td>16.4000000</td>
<td>9.1896483</td>
<td>0</td>
<td>37.0000000</td>
</tr>
<tr>
<td>NumBusStopsPerArea</td>
<td>NumBusStopsPerArea</td>
<td>36</td>
<td>21.8579133</td>
<td>17.489086</td>
<td>0</td>
<td>60.000000</td>
</tr>
<tr>
<td>NumGroceryStores</td>
<td>NumGroceryStores</td>
<td>36</td>
<td>2.0277777</td>
<td>2.1972205</td>
<td>0</td>
<td>12.000000</td>
</tr>
<tr>
<td>NumGroceryPer100HH</td>
<td>NumGroceryPer100HH</td>
<td>36</td>
<td>0.1959018</td>
<td>0.1965353</td>
<td>0</td>
<td>0.7052186</td>
</tr>
<tr>
<td>NumGroceryPer1000Pop</td>
<td>NumGroceryPer1000Pop</td>
<td>36</td>
<td>0.5349845</td>
<td>0.5026742</td>
<td>0</td>
<td>1.7406440</td>
</tr>
<tr>
<td>NumGroceryPerArea</td>
<td>NumGroceryPerArea</td>
<td>36</td>
<td>3.0361953</td>
<td>4.4416542</td>
<td>0</td>
<td>20.000000</td>
</tr>
<tr>
<td>Land Area</td>
<td>Land Area</td>
<td>36</td>
<td>1.2925000</td>
<td>1.3300798</td>
<td>0.1250000</td>
<td>7.1300000</td>
</tr>
<tr>
<td>TOTALPOP</td>
<td>TOTALPOP</td>
<td>36</td>
<td>3936.72</td>
<td>1901.33</td>
<td>116.00</td>
<td>9809.00</td>
</tr>
<tr>
<td>TotalHH</td>
<td>TotalHH</td>
<td>36</td>
<td>1214.07</td>
<td>608.2691856</td>
<td>317.000000</td>
<td>2784.00</td>
</tr>
<tr>
<td>PERCENT_LOWI</td>
<td>PERCENT_LOWI</td>
<td>36</td>
<td>15.0165656</td>
<td>14.0224616</td>
<td>0.5000000</td>
<td>67.7900000</td>
</tr>
</tbody>
</table>
3.4 Data Treatment

According to the number of Cook County census tracts deemed food deserts by the Food Access Research Atlas Data in 2010, I initially have 39 observations. However, due to outliers and inconsistencies between data sources, I am left with a total of 36 observations.

The outlier is the census tract that only contains O'Hare International Airport. With a total population of 26 people and no large grocery stores, I choose to remove the census tract from my study so it does not skew my results.

There are two instances where the food desert identified by the Research Atlas Data have actually combined with a non-food desert area into a new census tract on USBoundary.com. And so, the information available on USBoundary.com is only for the new combined census tract. Because I cannot locate the geographic center of the old food desert, nor obtain its land area, I will leave them out of my dataset.

3.5 Limitations

Although I have included nearly all of the population into my sample, based on available data, there are only 39 food deserts identified in Cook County. Because of the small number of observations used in my dataset, the results from my analysis may not be conclusive. However, they can at least indicate the general direction of the relationship between variables and suggest potential issues with the current definition of food deserts.

In addition, the geographic metrics are not perfect. The majority of census tracts are not symmetrical. Some census tracts have substantially large non-residential area, such as forest preserve, golf courses, and small lakes, that skew the average residential location away from the center. While I have used the geographic center of census tracts, just as previous studies have done, it is only a proxy.
4. Examining the Current State of Cook County Food Deserts Identified in 2010

4.1 Testing availability: Does the amount of low income residents affect the number of grocery stores in their census tract?

Previous research say that residents in lower income urban areas tend to face more barriers, particularly an inadequate number of grocery stores, that limit their access to healthy food. I want to see whether this holds true for food deserts in Cook County. Is there a negative relationship between the percentage of total population that is low income and the number of grocery stores within the census tract borders? And if so, what is the magnitude of the effect?

To find out, I first create a correlation matrix with various variables that represent grocery store availability:

1. Number of grocery stores in a food desert
2. Number of grocery stores per 100 households
3. Number of grocery stores per 1000 people in the food desert population
4. Number of grocery stores per land area

As we can see from Table 2, with a p-value of 0.0058, there is a significant positive correlation between the percentage of low income residents and the number of grocery stores per land area for food deserts in Cook County. This contrasts the statements made by previous literature.
To find an estimate of the magnitude of the effect of low income on grocery store availability in Cook County, I also conduct a weighted linear regression. Because the other variables do not have significant correlation with the percentage of total population that is low income, I use the number of grocery stores per land area as my response variable in my regression. I also incorporate population density as a weight to control for the difference in population and land area between food deserts.

Table 3. \( \text{NumGroceryPerArea} = \alpha w_1 + \beta_{w1} \text{PERCENT_LOWI} + \varepsilon \)

| Parameter Estimates |  |
|---------------------|--|---|------------------|-------------|---|-----|------|---|
| **Variable** | **Label** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** |
| Intercept | Intercept | 1 | 1.60260 | 1.33784 | 1.20 | 0.2392 |
| PERCENT_LOWI | PERCENT_LOWI | 1 | 0.17824 | 0.05388 | 3.31 | 0.0022 |

From Table 3, the p-value is 0.0022, and thus the effect of PERCENT_LOWI is significant. For every percentage increase of low income residents, the number of grocery stores per land area increases by 0.17824.
4.2 Testing Accessibility: Does the amount of low income residents affect the number of bus stops in their census tract?

Very few public transportation systems have actually designed their routes to provide convenient direct access to grocery stores for residents. Bus lines rarely are made to serve intra-neighborhood food shopping pattern. Rather, they are usually planned around commuter routes (Gottlieb, R). I want to see if there is a negative relationship between the percentage of total population that is low income in a food desert and the number of bus stops within the census tract borders.

To do so, I first create a correlation matrix with various variables that represent bus stop availability to see which ones were correlated with low income percentage:

1. Number of bus stops in a food desert
2. Number of bus stops per land area

As we can see from Table 4, with a p-value < 0.0001, there is a significant positive correlation between the percentage of low income residents and the number of bus stops per land area.

**Table 4. Correlation between Percentage of Low Income and Bus Stop Availability**

<table>
<thead>
<tr>
<th>Pearson Correlation Coefficients</th>
<th>Number of Observations</th>
<th>PERCENT_LOWI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob &gt;</td>
<td>r</td>
<td>under H0: Rho=0</td>
</tr>
<tr>
<td></td>
<td>NumBusStops</td>
<td>0.1645</td>
</tr>
<tr>
<td></td>
<td>NumBusPerArea</td>
<td>0.71298</td>
</tr>
<tr>
<td></td>
<td>NumBusPerArea</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Then, I use a weighted linear regression, with population density as the weight, to estimate the magnitude of the effect of the percentage of low income residents on bus stop availability. With a p-value that is near zero, I can conclude with 95% confidence that the
number of bus stops per area will increase by 0.90839 for every percent increase of low income residents in the total population. Again, these results pertaining to Cook County food deserts contrast what has been said in previous literature.

**Table 5.** $\text{NumBusPerArea} = \alpha_{w2} + \beta_{w2} \text{PERCENT_LOWI} + \epsilon$

<table>
<thead>
<tr>
<th>Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable</strong></td>
</tr>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td>PERCENT_LOWI</td>
</tr>
</tbody>
</table>

### 4.3 Testing Accessibility: Do residents in food deserts with a higher percentage of low income population have to walk farther to reach a grocery store?

For a final check to see if there are any other accessibility issues that I have not previously considered, I look at the relationship between the percentage of the total population that is low income and the walking distance from the center of the food desert to the nearest grocery store ($W\text{dist}_\text{CG}$). Again, I use population density as a weight in the linear regression.

With a p-value that is less than 0.10, I can conclude with 90% confidence that the percentage of low income residents has a significant effect on the walking distance from the center of the census tract to the nearest grocery store. In fact, for every percent increase in low income residents in the food desert population, the walking distance decreases by 0.01055 miles. Thus, the distance needed to carry grocery bags home from supermarkets decreases.

**Table 6.** $W\text{dist}_\text{CG} = \alpha_{w3} + \beta_{w3} \text{PERCENT_LOWI} + \epsilon$

<table>
<thead>
<tr>
<th>Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable</strong></td>
</tr>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td>PERCENT_LOWI</td>
</tr>
</tbody>
</table>
4.4 Summary of Results

My original expectation was that food deserts with higher percentage of low income residents will have lower accessibility to grocery stores and bus stops. However, with the positive correlation coefficients and positive beta coefficients, I discover that as of 2015, the food deserts in Cook County with higher percentage of low income residents actually have more access to both grocery stores and bus stops. The result of regressing PERCENT_LOWI and Wdist_CG show that the walking distance decreases as the percentage of low income residents increase, which also supports this finding.

The relationship between the percentage of low income residents and the availability of grocery stores and bus stops in Cook County food deserts are directionally correct, contrary to the observations of previous food desert areas studies. How can this be?

5. Assessing the Disparity of Different Food Desert Levels

Perhaps there are different severity levels of food deserts. Maybe some food deserts have lower access to grocery stores than others. And on the other side, maybe some food deserts have better access to grocery stores. These "less severe" food deserts may be what are causing the positive correlation between low income and accessibility. Unfortunately, I am not able to find previous literature that classifies food deserts by severity, and so I have to create my own categorization.

Based off of the USDA definition, using the walking distance from the geographic center of the food desert to the nearest grocery store, I divide the food deserts into four groups. (See Figure 4).

Group 1 (in red) has a walking distance ≥ 1 mile. Census tracts in this group are "true" food deserts because they meet the USDA requirement that the walking distance to the nearest
grocery store is greater than 1 mile. Even within this group of the highest severity, we can see there is a vast difference between the distances, with a range from 1.1 miles to 3.2 miles to the nearest grocery store. These food deserts are the ones that, just looking at the distance, require the most attention.

Group 2 (in orange) has a walking distance $\epsilon [0.5, 1.0]$ mile. Food deserts that fall in Group 2 are not technically food deserts, as they are under the USDA distance threshold. However, half a mile is still a considerable amount to walk while carrying groceries, and so, Group 2 food deserts can also use assistance.

Group 3 (in yellow) has a walking distance $\epsilon [0.25, 0.5]$ mile, and Group 4 (in green) has a walking distance $\epsilon [0, 0.25]$ mile. These food deserts, while identified in 2010, may no longer be of concern in terms of accessibility issues, as improvements seem to have been made in the past five years.

**Figure 4.** Grouping Food Deserts By Walking Distance to Nearest Grocery Store

<table>
<thead>
<tr>
<th>Census Tract</th>
<th>Walk Dist C6</th>
<th>Dist $\geq$ 1</th>
<th>Dist $\epsilon $</th>
<th>Dist $\epsilon [0.5, 1]$</th>
<th>Dist $\epsilon [0.25, 0.5]$</th>
<th>Dist $\epsilon [0, 0.25]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1070191701</td>
<td>2.7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701930001</td>
<td>2.7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701927000</td>
<td>1.9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701926900</td>
<td>1.7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701924900</td>
<td>1.4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701929000</td>
<td>1.2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701927700</td>
<td>1.1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701923700</td>
<td>1.1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701920100</td>
<td>1.1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701950200</td>
<td>0.9</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701925700</td>
<td>0.9</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701950300</td>
<td>0.9</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701950400</td>
<td>0.85</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701924700</td>
<td>0.8</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701924400</td>
<td>0.9</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701927500</td>
<td>0.7</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701950600</td>
<td>0.7</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701950800</td>
<td>0.7</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701950900</td>
<td>0.6</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701923500</td>
<td>0.6</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701922500</td>
<td>0.5</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701912200</td>
<td>0.5</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701922800</td>
<td>0.5</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701922900</td>
<td>0.4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701925901</td>
<td>0.4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701927700</td>
<td>0.3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701928500</td>
<td>0.3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701929300</td>
<td>0.2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701939000</td>
<td>0.2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701950100</td>
<td>0.2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701950000</td>
<td>0.2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701950010</td>
<td>0.2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701925901</td>
<td>0.2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701929400</td>
<td>0.2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10701931000</td>
<td>0.2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
To further illustrate the difference in distribution of walking distance from the center to the nearest grocery store between groups, I use the distribution of the Wilcoxon scores. Wilcoxon scores are, simply put, ranks of the observations to see if the two samples are from the same distribution. With Figure 5, we can see that there is a significant difference between the four groups' walking distance distribution.

**Figure 5.** Confirmation of the Difference in Walking Distance between Groups

5.1 Testing Accessibility Based on Food Desert Severity Levels Using Number of Grocery Stores Per Land Area

After dividing the food deserts into severity levels by walking distance, I also want to see if there is a significant difference in the number of grocery stores available per land area between groups.

Because the data may not be normally distributed and the groups of food deserts are independent, I need to use a non-parametric test. Assuming individual groups are random
samples, I use the Wilcoxon Mann-Whitney U-test to see if the two groups are from the same distribution or not, based on the location parameters, such as the median. In particular, to visually compare the distribution of walking distance, I first look at the distribution of the Wilcoxon rank sum scores between the four groups.

**Table 7.** Wilcoxon rank sums to compare grocery store availability between food desert groups

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Sum of Scores</th>
<th>Expected Under H0</th>
<th>Std Dev Under H0</th>
<th>Mean Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>82.0</td>
<td>185.0</td>
<td>28.211503</td>
<td>8.200000</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>251.0</td>
<td>259.0</td>
<td>30.705410</td>
<td>17.928571</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>83.0</td>
<td>74.0</td>
<td>19.794500</td>
<td>20.750000</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>250.0</td>
<td>148.0</td>
<td>26.185662</td>
<td>31.250000</td>
</tr>
</tbody>
</table>

Average scores were used for ties.

**Figure 6.** Wilcoxon Rank Sums to compare grocery store availability between food desert groups
As we can see from the box plot in Figure 6, there is a significant amount of overlap between Group 2 and 3. The mean Wilcoxon scores of the two groups, 17.928 and 20.750, are close together. On the other hand, Group 1 and 2 seem to have a clear difference in mean scores, as well as Group 3 and Group 4 (Table 7).

To further test if there exists differences in distribution between groups, as mentioned above, I conduct a non-parametric two-sample t test, also known as the Wilcoxon Mann-Whiney U-Test (WMW test) between consecutive groups. My hypothesis is:

H₀: median NumGroceryPerArea for Group i < median NumGroceryPerArea for Group j, where i<j vs. Hₐ: r>H₀

Table 8. Nonparametric two-sample t test to compare NumGroceryPerArea between consecutive food desert severity groups

<table>
<thead>
<tr>
<th>Group 1 vs Group 2</th>
<th>Group 2 vs Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wilcoxon Two-Sample Test</strong></td>
<td><strong>Wilcoxon Two-Sample Test</strong></td>
</tr>
<tr>
<td><strong>Statistic</strong></td>
<td>80.0000</td>
</tr>
<tr>
<td><strong>Normal Approximation</strong></td>
<td></td>
</tr>
<tr>
<td>Z</td>
<td>2.6380</td>
</tr>
<tr>
<td>One-Sided Pr &lt; Z</td>
<td>0.0042</td>
</tr>
<tr>
<td>Two-Sided Pr &gt;</td>
<td>0.0083</td>
</tr>
<tr>
<td><strong>t Approximation</strong></td>
<td></td>
</tr>
<tr>
<td>One-Sided Pr &lt; Z</td>
<td>0.0074</td>
</tr>
<tr>
<td>Two-Sided Pr &gt;</td>
<td>0.0147</td>
</tr>
</tbody>
</table>

Z includes a continuity correction of 0.5.

Page 20 of 37
According to the one-sided p-values that are less than $\alpha = 0.05$, there are significant differences between the median number of grocery store per area for Groups 1 and 2, and for Groups 3 and 4. However, with a large p-value of 0.2496, I have to reject the null hypothesis for Group 2 and 3 and conclude that there is not a significant difference in the number of grocery stores per area for the two groups.

5.2 Testing Accessibility Based on Food Desert Severity Levels Using Number Grocery Stores Per 100 Households

The one flaw of using the PerArea control is that even though it controls for the discrepancy in land area between food deserts, it does not take into account the number of residents, which can affect results. For example, two food deserts may have the same land area and the same number of grocery stores per land area. But a golf course takes up a majority of food desert A. And so, in comparison, there are fewer people in food desert A than in B because there is less residential land. In addition, food desert A residents are more densely packed and are closer to the nearest grocery store than residents in food desert B, who are more equally spread out over the census tract. Even though food deserts A and B have the level of severity based off of number of grocery stores per area, they do in fact have different levels of accessibility once population is taken into account.

To mitigate this problem, I conduct a WMW test using the median number of grocery store per 100 households as my test statistic.

$H_0$: median NumGroceryPer100HH for Group $i <$ median NumGroceryPer100HH for Group $j$, where $i<j$ vs. $H_A: r-H_0$
Table 9. Nonparametric two-sample t test to compare NumGroceryPer100HH between consecutive food desert severity groups

<table>
<thead>
<tr>
<th></th>
<th>Group 1 vs Group 2</th>
<th>Group 2 vs Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Statistic</strong></td>
<td>98.0000</td>
<td>39.0000</td>
</tr>
<tr>
<td><strong>Normal Approximation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Z</td>
<td>-1.5709</td>
<td>0.0531</td>
</tr>
<tr>
<td>One-Sided Pr &lt; Z</td>
<td>0.0581</td>
<td>0.4788</td>
</tr>
<tr>
<td>Two-Sided Pr &gt;</td>
<td>0.1162</td>
<td>0.9576</td>
</tr>
<tr>
<td>Z includes a continuity correction of 0.5.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**t Approximation**

<table>
<thead>
<tr>
<th></th>
<th>Group 3 vs Group 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Statistic</strong></td>
<td>16.0000</td>
</tr>
<tr>
<td><strong>Normal Approximation</strong></td>
<td></td>
</tr>
<tr>
<td>Z</td>
<td>-1.6135</td>
</tr>
<tr>
<td>One-Sided Pr &lt; Z</td>
<td>0.0533</td>
</tr>
<tr>
<td>Two-Sided Pr &gt;</td>
<td>0.1066</td>
</tr>
<tr>
<td>Z includes a continuity correction of 0.5.</td>
<td></td>
</tr>
</tbody>
</table>

The one-sided p-values that are less than $\alpha = 0.10$ for the tests between Groups 1 and 2 and Groups 3 and 4. Thus I can say with 90% confidence that there are significant differences between the median number of grocery store per 100 households for between these two sets of groups. However, just like with the test using the number of grocery stores per area, I have to reject the null hypothesis for Group 2 and 3 and conclude that there is not a significant difference in the number of grocery stores per 100 households for the two groups.
In addition, I also want to test if there was a significant difference in grocery store availability between the Group 1 food deserts that meet the USDA low access definition and the rest (Groups 2, 3, and 4) that do not. To do so, I combine Groups 2, 3, and 4 into one sample and conduct another Wilcoxon U-test between Group 1 and the combined Groups 2, 3, and 4. The test statistic remains the median number of grocery stores per 100 households.

H₀: median NumGroceryPer100HH for Group 1 < median NumGroceryPer100HH for Group 2, 3, and 4 vs. Hₐ: −H₀

Table 10. Nonparametric two-sample t test to compare NumGroceryPer100HH between Group 1 and combined Groups 2, 3, and 4

<table>
<thead>
<tr>
<th>Wilcoxon Two-Sample Test</th>
<th>Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal Approximation</td>
<td></td>
</tr>
<tr>
<td>Z</td>
<td>-2.2863</td>
</tr>
<tr>
<td>One-Sided Pr &lt; Z</td>
<td>0.0111</td>
</tr>
<tr>
<td>Two-Sided Pr &gt;</td>
<td>Z</td>
</tr>
<tr>
<td>t Approximation</td>
<td></td>
</tr>
<tr>
<td>One-Sided Pr &lt; Z</td>
<td>0.0142</td>
</tr>
<tr>
<td>Two-Sided Pr &gt;</td>
<td>Z</td>
</tr>
</tbody>
</table>

Z includes a continuity correction of 0.5.

In Table 10, we can see that the p-value = 0.0142, which makes the difference in the median grocery store availability per 100 households between Group 1 and the other groups significant at a 98% confidence level.

5.3 Summarizing the Results

From the Wilcoxon Mann-Whiney U-Tests, I conclude that Group 1 and Group 4 are significantly different from Groups 2 and 3. This indicates that rather than breaking the food deserts down into four different groups via walking distance from the center to the nearest
grocery store, it is better to combine Group 2 and 3 into one level of severity due to the similar distribution in grocery store availability.

Food deserts in Group 1 require the most immediate attention, due to them having the farthest walking distance from center to nearest grocery store, and the lowest number of grocery stores available in proportion to land area and number of households. This means that 29% of identified food deserts in Cook County, which houses 25.3% of the total population of Cook County residents in food deserts, are highly severe and must be treated with the highest priority.

However, with the little difference between Groups 2 and 3, contrary to my original thought, I believe that both groups, not just Group 2, still need assistance in improving grocery store accessibility.

6. Reexamining the Definition of Food Desert

The classification created in the previous section relies on the current USDA definition of low access as being more than 1 mile walking distance from a supermarket or large grocery store. But what happens if I incorporate public transportation into the definition and use the walking distance after using the bus to access the nearest grocery store as the new criteria? Will there be changes in the levels of accessibility for the four levels of severity?

To iterate, below (Figure 7) is an example of how much distance can be saved by using public transportation instead of walking. Looking at the numbers boxed in red, we can see that the walking distance from the geographic center to the nearest grocery store without taking the bus is 0.6 miles. This measurement puts the food desert in Group 2 severity (walking distance = 0.6 ∈ [0.5, 1.0] mile), which means that the tract requires a moderate amount of attention from Cook County to alleviate accessibility issues. However, the total required walking distance after taking the bus is a little over 0.2 miles, with 0.2 miles needed to walk to the bus stop, and 308
feet to walk from the bus stop to the grocery store. This new distance re-categorizes the food desert as Group 4 (walking distance $= 0.2 \epsilon [0, 0.25]$ mile), a low priority because accessibility thanks to public transportation is not much of a problem in comparison to other food deserts. 

**Figure 7.** Walking distance to nearest grocery store with and without use of public transportation

Expanding this to all 36 food deserts in my sample, I have re-classified and re-prioritized food deserts based on the walking distance to and from the appropriate bus stops to reach the nearest grocery store.

If one can assume that everyone can take public transportation, by incorporating buses, I can isolate and identify the census tracts that are the most severe food deserts and those that are no longer food deserts. As shown by Figure 8, I have reduced the percentage of food deserts in Group 1 from 29% of the total number of food deserts in Cook County to just 14%. This saves
both time and resources, as the County and the City of Chicago only need to focus their immediate attention on five food deserts of the highest severity instead of ten.

In addition, Group 2 food deserts have already decreased by half, while Group 3 and Group 4 have increased. Thus, provided there is universal access to public buses, 34% of the original 36 Cook County food deserts in my sample no longer need to be identified as food deserts, because they do not have an accessibility issue to grocery stores with the use of public transportation.

**Figure 8.** Impact of Public Transportation on Classification of Cook County Food Deserts

6.1 Identification of the Five Most Severe Food Deserts in Cook County, IL

In the order of highest severity based on walking distance between center and nearest grocery store after using public transportation, the five food deserts are:

1. Census tract 8300.01 / Timber Ridge Rd in Matteson, IL
2. Census tract 8117.01 / Westdale Park St in Franklin Park / Bartlett, IL
3. Census tract 8258.01 / Downey Park in Calumet City / Dolton, IL
4. Census tract 8257 / Burnham Woods Golf Course St in Calumet City, IL
5. Census tract 4201 / West Lagoon St in Woodlawn, Chicago, IL

7. Assessing Possible Transportation-Related Solutions for Cook County Food Deserts

So what can be done to improve the accessibility and availability problems, especially for the Group 1 food deserts that are the most severe?

7.1 Factors of Consideration

As of 2013, both the City of Chicago as well as Illinois lawmakers have focused on programs to create more brick and mortar grocery stores and farmers markets. However, not much research has been done on transportation-related strategies. A number of questions need to be considered to compare which strategy is the most effective for Cook County food desert residents.

1. Will this initiative actually increase accessibility? And by how much?
2. Who will operate and fund this project?
3. How much are the startup costs? How much are yearly operation costs?
4. Is there some pre-existing form of this service?
5. Will employees be needed? If so, how many?
6. Are there any policy changes required?
7. Are there any limitations? What potential problems may arise and how can they be combated or prevented?

Additionally, values and preferences for low-income, food desert residents must also be taken into account. The most important factor is that groceries, and subsequently accessing these
groceries, must be low cost. Secondly, getting groceries must be convenient in both time and distance. Ideally, the facility should have a wide produce selection, as well as flexible payment options, such as the acceptance of food stamps. And lastly, food desert residents should be able to have continuous, year-round access to buying groceries.

Below is a general comparison of various transportation-related solutions. To have the biggest impact on using our limited resources, I encourage policymakers to first target these solutions towards combating the food accessibility problems in the five most severe food deserts identified as Group 1.

**Table 11.** Transportation-Related Options to Improve Healthy Food Accessibility

<table>
<thead>
<tr>
<th></th>
<th>Brings Food to People</th>
<th>Brings People to Food</th>
<th>Additional Costs To Consumer</th>
<th>Operates Year-Round</th>
<th>Reduced Food Price</th>
<th>Prepay for Food</th>
<th>Wide Selection of Produce Available</th>
<th>No Time Constraint on Accessibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grocery-Run Shuttle</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Modify Current Routes</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Supermarket Shuttle Service</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Mobile Produce Trucks</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Mobile Farmers Market</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Virtual Supermarket</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
7.2 Analyzing Possible Solutions

7.2.1. Grocery Store-Run Shuttle

The option that interests me the most is the grocery store-run shuttle. It is an incentive-based shuttle service that is operated by a large grocer, and provides free or low-cost transportation for shoppers from the store (Swingley, S.). The benefits of this shuttle is that is more cost-effective for people and saves them the hassle of walking with heavy bags of groceries to the nearest bus stop, then carrying them on the bus. On the business side, grocery store-run shuttles can improve customer loyalty and draw in new shoppers. The grocer can start with one van as a trial period and increase the scale as demand and awareness increases. Unfortunately, the drawback is the initial cost for buying a vehicle, hiring employees, and potentially marketing the service.

However, a study, "Doing Well by Doing Good? A Supermarket Shuttle Feasibility Study," conducted in 2002 by Diana Cassady and Vidhya Mohan on the profitability of supermarket-run shuttles, proved that these costs will not be a problem. The shuttles in the study operated from 8am to 9pm, and offered customers a free ride home with a $25 purchase. Within two to five months of buying a 15-passenger van, assuming that 10% of households in the area used the shuttle, the stores were able to break even. For one particular small chain grocer, the shuttle service costs, such as drivers, operating costs, and maintenance was only 1% of the gross store revenues.

If there seems to be a reluctance to start such programs, Cook County and the City of Chicago can even incentivize supermarkets and larger grocery stores to create these shuttle services by subsidizing the initial costs or offering tax breaks on the profits made through the shuttles.
7.2.2. Modify Public Transit

Being able to reduce the number of severe food deserts in Group 1 is contingent on everyone in Cook County food deserts having better access to public transportation and are more enticed to it. Currently, the problem with most food deserts’ buses is that bus stops are only along the perimeter of the census tract. Increasing bus stops to drive inside the census tract and increasing the frequency of routes will make using public transportation more convenient. This approach reduces the distance residents need to walk to get to the nearest bus stop, as well as reduce the amount of time they need to wait for the next bus. However, the method of increasing bus stops in general and not just targeting grocery stores is not the most cost efficient. Modifying public transit will require a lot of funding from the City of Chicago and Cook County for higher operational and labor costs.

7.2.3. Supermarket Shuttle Service

Unlike just changing the regular public transportation routes, supermarket shuttles have a direct service to connect people to grocers. These "Stop & Shop" shuttles operate on a weekly basis and drive riders to various food locations (Swingley, S). The best part about supermarket shuttles is that they are usually free for the rider, unlike public transportation. But these shuttles cannot make as many stops to pick up residents as public buses. And thus, people will likely have to walk a farther distance to access the shuttle. In addition, starting these shuttles, hiring employees, and designing efficient bus routes will be costly and time-consuming. While this solution is a good mid-term or long-term option, it is not something Cook County can immediately implement to see improvements.
7.2.4. Mobile Produce Trucks

More and more states are starting to use vans or buses to bring healthy food directly to people in low-income areas. Mobile produce trucks do not require as much walking for residents as brick and mortar stores do, and still offer a wide enough range of healthy produce that tend to be locally grown. They also operate year-round and can make multiple stops in a day to different areas. Unfortunately, the start-up and operational costs of running a mobile produce truck is extremely high. Profits also tend to be low to provide people in low-income communities a reasonable price for groceries. A large, consistent amount of government funding and/or donations will be required to support this service.

7.2.5. Mobile Farmers Market

The City of Chicago has focused a lot of energy on launching farmers markets in food deserts. Farmers markets are great because they carry locally-grown produce. When residents buy food from farmers markets, their money stays closer to the community rather than getting set to large chain retailers and distributors. Mobile farmers markets are one step better, as they deliver this locally-grown food to residents rather than making people travel far distances to access a farmers market location. However, farmers markets and mobile farmers markets alike are impractical because only operate during the growing season, typically between April and October (Swingley, S). For the other half of the year, residents of food deserts need to find somewhere else to buy their groceries from, which brings them back to square one. While it is great that the City is so supportive of local products, there are more effective solutions to combat food deserts’ low access problem.
7.2.6. Virtual Supermarket

The virtual supermarket is incredibly convenient as it brings groceries to people's homes and offers access to a full-scale grocery store selection. However, this concept is not the most practical for low income families. First of all, not all low income families may have access to a computer or internet. Second, even though the delivery service provides convenience and saves time, the main priority of low income households is to keep expenditure costs low. The surcharge, even if subsidized, is a significant portion of money that cannot be overlooked. If there is no delivery charge, food is reasonably priced, and internet access was made available at local schools or libraries, this may be a good option for Cook County. But this will require a good amount of funding or fundraising and time. So as of now, the costs of a virtual supermarket delivery outweigh the benefits.

8. Conclusion

From reassessing the Cook County food deserts identified using the USDA definition, I have come to the conclusion that not all food deserts are equal. Previous studies only classify census tracts as food deserts but do not categorize these food deserts into levels of severity. Some census tracts have been incorrectly classified as food deserts, or should no longer be considered as such thanks to improvements made between the time of identification and present day. Based on the new definition incorporating the effect of public transportation on walking distance, providing that all residents have access to buses, I have identified the five most severe food deserts in Cook County. The city and county can make the best use of their limited resources by first focusing on improving the accessibility problem in these five food deserts.

Transportation-related options are good starting points, but they will not be able to completely solve the food desert problem we see in Cook County. There are a multitude of other
factors that need to be tackled, such as the lack of educational opportunities, inflated prices, and limited selections of low-quality produce. With the increased awareness of food deserts in recent years, action has been taken to help the situation. Slowly but surely, by first targeting the most severe food deserts and then increasing our focus to less severe areas, we can fight against the lack of accessibility to healthy food in the local communities in Cook County and expand to help the nation.

9. Opportunities for Further Research

If provided the appropriate access to information and contacts, I would love to conduct a quantitative cost-benefit analysis of each transportation-related solution. However, many different factors that are beyond the scope of this thesis need to be considered in order to do so.

Reaching the supermarket is only half the battle against food deserts. The other half occurs within the store. Ideally, I would like to find or collect data on the prices, quality, and sale quantity of fresh produce provided in the nearest grocery store for each identified food desert in Cook County. Then, using this information, I want to see if there are any trends between the average income or disposable income of the food desert residents, the price and quality of the produce available to them, and their purchase behavior.

With the same data, I would like to compare the quality of produce between food deserts to see if there are significant enough differences to classify food deserts by food quality as well.
References


Appendix

Sample of SAS Codes Used

proc import out=work.fd_censustracts datafile='/sscc/home/b/bxj628/Thesis/Cook County Food Deserts 1.xlsx' dbms=xlsx replace;
   sheet="SAS``;
   getnames=YES;
run;

1. Correlation Matrix

/*Which Grocery Store Availability Variables are correlated with PERCENT_LOWI? */
proc corr data=work.fd_censustracts;
   VAR PERCENT_LOWI;
   WITH NumGroceryStores NumGroceryPer100HH NumGroceryPer1000Pop NumGroceryPerArea;
run;

/*Which Public Transportation Availability Variables are correlated with PERCENT_LOWI?*/
proc corr data=work.fd_censustracts;
   VAR PERCENT_LOWI;
   WITH NumBusStops NumBusPerArea;
run;

2. Weight Linear Regression

/*Weighted Regression b/t %Low Income and Grocery Stores per Area*/
Data work.fd_censustracts;
Set work.fd_censustracts;
   PopDensity = TOTALPOP/Land_Area;
Run;

proc reg data=work.fd_censustracts;
   model NumGroceryPerArea = PERCENT_LOWI;
   weight PopDensity;
run;

/*Weighted Regression b/t %Low Income and Transportation*/
proc reg data=work.fd_censustracts;
   model NumBusPerArea = PERCENT_LOWI;
   weight PopDensity;
run;

3. Wilcoxon Mann-Whitney U-Test
/*Grocery Store Per Area*/
/*Rank Sum Scores Distribution*/
proc npar1way wilcoxon data=work.fd_censustracts;
   title "Nonparametric test to compare grocery store availability between food desert groups";
   class group;
   var NumGroceryPerArea;
run;

/*Group 1 and 2*/
proc NPAR1WAY data=work.fd_censustracts wilcoxon;
   title "Nonparametric test to compare grocery store availability between food desert groups";
   where Group <= 2;
   class group;
   var NumGroceryPerArea;
run;

/*Group 2 and 3*/
proc NPAR1WAY data=work.fd_censustracts wilcoxon;
   title "Nonparametric test to compare grocery store availability between food desert groups";
   where Group >= 3, <=2;
   class group;
   var NumGroceryPerArea;
run;

/*Group 3 and 4*/
proc NPAR1WAY data=work.fd_censustracts wilcoxon;
   title "Nonparametric test to compare grocery store availability between food desert groups";
   where Group >= 3;
   class group;
   var NumGroceryPerArea;
run;

/*****************************/
/*Distance CG*/
proc NPAR1WAY data=work.fd_censustracts wilcoxon;
   title "Nonparametric test to compare grocery store availability between food desert groups";
class group;
var Wdist_CG;
run;

/*****************************/
/*/Grocery Store Per 100HH*/
proc NPAR1WAY data=work.fd_censustracts wilcoxon;
   title "Nonparametric test to compare grocery store availability between food desert groups";
   class group;
   var NumGroceryPer100HH;
run;

/*Group 1 and 2*/
proc NPAR1WAY data=work.fd_censustracts wilcoxon;
   title "Nonparametric test to compare grocery store availability between food desert groups";
   where Group <= 2;
   class group;
   var NumGroceryPer100HH;
run;

/*Group 2 and 3*/
proc NPAR1WAY data=work.fd_censustracts wilcoxon;
   title "Nonparametric test to compare grocery store availability between food desert groups";
   where Group >= 3, <= 2;
   class group;
   var NumGroceryPer100HH;
run;

/*Group 3 and 4*/
proc NPAR1WAY data=work.fd_censustracts wilcoxon;
   title "Nonparametric test to compare grocery store availability between food desert groups";
   where Group >= 3;
   class group;
   var NumGroceryPer100HH;
run;